#2

W/573317

PCT/GB 89/00460

- 2 May 1989

The Patent Office

THE PATENT OFFICE
STATE HOUSE
66—71 HIGH HOLBORN
LONDON WC1R 4TP

07 Rec'd PCT/PTO 14 DEC 1989

08/239012

REC'D 17 JUL 1989
WIPO PCT

I, the undersigned, being an officer duly authorised in accordance with Section 62(3) of the Patents and Designs Act 1907, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or the inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words, "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Witness my hand this
9th day of JUNE 1989.

M. Lunell

PRIORITY DOCUMENT

PATENTS ACT 1977

- 3 MAY 88

PATENTS FORM NO. 1/77 (Revised 1982)

(Rules 16, 19)

The Comptroller
The Patent Office

£10

REQUEST FOR GRANT OF A PATENT

8810400.5

THE GRANT OF A PATENT IS REQUESTED BY THE UNDERSIGNED ON THE BASIS OF THE PRESENT APPLICATION

I Applicant's or Agent's reference *(Please insert if available)* PP/7644

II Title of invention

Analysing Polynucleotide Sequences

III Applicant or Applicants *(See note 2)*

Name (First or only applicant) Edwin Southern

Country England State ADP Code No

Address Department of Biochemistry, University of Oxford,
South Parks Road, Oxford OX1 3QU, England.

Name (of second applicant, if more than one)

Country State

Address

IV Inventor (see note 3)

(a) The applicant(s) is/~~are~~ the sole/~~joint~~ inventor(s)

~~or~~

~~(b) A statement accompanies this Patents Form~~
~~No 7/77 is/will be furnished~~

V Name of Agent (if any) *(See note 4)* Stevens, Hewlett & Perkins ADP CODE NO

VI Address for Service *(See note 5)*

5 Quality Court
Chancery Lane
London WC2A 1HZ

VII Declaration of Priority *(See note 6)*

Country	Filing date	File number

IX Check List *(To be filled in by applicant or agent)*

A The application contains the following number of sheet(s)

1 Request .....1..... Sheet(s)

2 Description .....14..... Sheet(s)

3 Claim(s) .....2..... Sheet(s)

4 Drawing(s) ..... Sheet(s)

5 Abstract ..... Sheet(s)

B The application as filed is accompanied by:-

1 Priority document .

Translation of priority document

3 Request for Search .....

4 Statement of Inventorship and Right to Grant .....

X It is suggested that Figure No.....of the drawings (if any) should accompany the abstract when published.

XI Signature *(See note 8)* Stevens Hewlett & Perkins

NOTES:

1. This form, when completed, should be brought or sent to the Patent Office together with the prescribed fee and two copies of the description of the invention, and of any drawings.

2. Enter the name and address of each applicant. Names of individuals should be indicated in full and the surname or family name should be underlined. The names of all partners in a firm must be given in full. Bodies corporate should be designated by their corporate name and the country of incorporation and, where appropriate, the state of incorporation within that country should be entered where provided. Full corporate details, eg a "corporation organised and existing under the laws of the State of Delaware, United States of America", trading styles, eg "trading as xyz company", nationality, and former names, eg "formerly (known as) ABC Ltd" are *not* required and should *not* be given. Also enter applicant(s) ADP Code No.(if known).

3. Where the applicant or applicants is/are the sole inventor or the joint inventors, the declaration (a) to that effect at IV should be completed, and the alternative statement (b) deleted. If, however, this is not the case the declaration (a) should be struck out and a statement will then be required to be filed upon Patent Form No 7/77.

4. If the applicant has appointed an agent to act on his behalf, the agent's name and the address of his place of business should be indicated in the spaces available at V and VI. Also insert agent's ADP Code No. (if known) in the box provided.

5. An address for service in the United Kingdom to which all documents may be sent must be stated at VI. It is recommended that a telephone number be provided if an agent is not appointed.

6. The declaration of priority at VII should state the date of the previous filing and the country in which it was made and indicate the file number, if available.

7. When an application is made by virtue of section 8(3), 12(6), 15(4) the appropriate section should be identified at VIII and the number of the earlier application or any patent granted thereon identified.

8. Attention is directed to rules 90 and 106 of the Patent Rules 1982.

9. Attention of applicants is drawn to the desirability of avoiding publication of inventions relating to any article, material or device intended or adapted for use in war (Official Secrets Acts, 1911 and 1920). In addition after an application for a patent has been filed at the Patent Office the comptroller will consider whether publication or communication of the invention should be prohibited or restricted under section 22 of the Act and will inform the applicant if such prohibition is necessary.

10. Applicants resident in the United Kingdom are also reminded that, under the provisions of section 23 applications may not be filed abroad without written permission or unless an application has been filed not less than

ANALYSING POLYNUCLEOTIDE SEQUENCES

1. INTRODUCTION

Three methods dominate molecular analysis of nucleic acid sequences: gel electrophoresis of restriction fragments, molecular hybridisation, and the rapid DNA sequencing methods. These three methods have a very wide range of applications in biology, both in basic studies, and in the applied areas of the subject such as medicine and agriculture. Some idea of the scale on which the methods are now used is given by the rate of accumulation of DNA sequences, which is now well over one million base pairs a year. However, powerful as they are, they have their limitations. The restriction fragment and hybridisation methods give a coarse analysis of an extensive region, but are rapid; sequence analysis gives the ultimate resolution, but it is slow, analysing only a short stretch at a time. There is a need for methods which are faster than the present methods, and in particular for methods which cover a large amount of sequence in each analysis.

This invention provides a new approach which produces both a fingerprint and a partial or complete sequence in a single analysis, and may be used directly with complex DNAs and populations of RNA without the need for cloning.

In one aspect the invention provides apparatus for analysing a polynucleotide sequence, comprising a support and attached to a surface therof an array of the whole or a chosen part of a complete set of oligonucleotides of a chosen length, the oligonucleotides being capable of taking part in hybridisation reactions. For studying mutations in a known polynucleotide sequence, the invention provides in another aspect apparatus comprising a support and attached to a surface therof an array of the whole or a chosen part of a complete set of oligonucleotides of a chosen

length comprising the known polynucleotide sequence, the oligonucleotides being capable of taking part in hybridisation reactions.

In another aspect, the invention provides a method of analysing a polynucleotide sequence, by the use of a support to the surface of which is attached an array of the whole or a chosen part of a complete set of oligonucleotides of a chosen length, which method comprises labelling the polynucleotide sequence or fragments thereof to form labelled material, applying the labelled material under hybridisation conditions to the array, and observing the location on the surface of the label associated with particular members of the set of oligonucleotides.

The idea of the invention is thus to provide a structured array of the whole or a chosen part of a complete set of oligonucleotides of a chosen length. The array, which may be laid out on a supporting film or glass plate, forms the target for a hybridisation reaction. The chosen conditions of hybridisation and the length of the oligonucleotides must at all events be sufficient for the available equipment to be able to discriminate between exactly matched and mismatched oligonucleotides. In the hybridisation reaction, the array is explored by a labelled probe, which may comprise oligomers of the chosen length or longer polynucleotide sequences or fragments, and whose nature depends on the particular application. For example, the probe may comprise labelled sequences amplified from genomic DNA by the polymerase chain reaction, or a mRNA population, or a complete set of oligonucleotides from a complex sequence such as an entire genome. The end result is a set of filled cells corresponding to the oligonucleotides present in the analysed sequence, and a set of "empty" sites corresponding to the sequences

which are absent in the analysed sequence. The pattern produces a fingerprint representing all of the sequence analysed. In addition, it is possible to assemble most or all of the sequence analysed if an oligonucleotide length is chosen such that most or all oligonucleotide sequences occur only once.

The number, the length and the sequences of the oligonucleotides present in the array "lookup table" also depend on the application. The array may include all possible oligonucleotides of the chosen length, as would be required if there was no sequence information on the sequence to be analysed. In this case, the preferred length of oligonucleotide used depends on the length of the sequence to be analysed, and is such that there is likely to be only one copy of any particular oligomer in the sequence to be analysed. Such arrays are large. If there is any information available on the sequence to be analysed, the array may be a selected subset. For the analysis of a sequence which is known, the size of the array is of the same order as length of the sequence, and for many applications, such as the analysis of a gene for mutations, it can be quite small. These factors are discussed in detail in what follows.

2. OLIGONUCLEOTIDES AS SEQUENCE PROBES

Oligonucleotides form base paired duplexes with oligonucleotides which have the complementary base sequence. The stability of the duplex is dependent on the length of the oligonucleotides and on base composition. Effects of base composition on duplex stability can be greatly reduced by the presence of high concentrations of quaternary or tertiary amines. However, there is a strong effect of mismatches in the oligonucleotides duplex on the thermal stability of the hybrid, and it is this which makes the technique of

hybridisation with oligonucleotides such a powerful method for the analysis of mutations, and for the selection of specific sequences for amplification by DNA polymerase chain reaction. The position of the mismatch affects the degree of destabilisation. Mismatches in the centre of the duplex may cause a lowering of the Tm by 10°C compared with 1°C for a terminal mismatch. There is then a range of discriminating power depending on the position of mismatch, which has implications for the method described here. There are ways of improving the discriminating power, for example by carrying out hybridisation close to the Tm of the duplex to reduce the rate of formation of mismatched duplexes, and by increasing the length of oligonucleotide beyond what is required for unique representation. A way of doing this systematically is discussed.

3. ANALYSIS OF A PREDETERMINED SEQUENCE

One of the most powerful uses of oligonucleotide probes has been in the detection of single base changes in human genes. The first example was the detection of the single base change in the betaglobin gene which leads to sickle cell disease. There is a need to extend this approach to genes in which there may be a number of different mutations leading to the same phenotype, for example the DMD gene and the HPRT gene, and to find an efficient way of scanning the human genome for mutations in regions which have been shown by linkage analysis to contain a disease locus for example Huntington's disease and Cystic Fibrosis. Any known sequence can be represented completely as a set of overlapping oligonucleotides. The size of the set is $N \quad s + 1 \eqsim N$, where N is the length of the sequence and s is the length of an oligomer. A gene of 1 kb for example, may be divided into an overlapping set of

around one thousand oligonucleotides of any chosen length. An array constructed with each of these oligonucleotides in a separate cell can be used as a multiple hybridisation probe to examine the homologous sequence in any context, a single-copy gene in the human genome or a messenger RNA among a mixed RNA population, for example. The length s may be chosen such that there is only a small probability that any oligomer in the sequence is represented elsewhere in the sequence to be analysed. This can be estimated from the expression given in the section discussing statistics below. For a less complete analysis it would be possible to reduce the size of the array e.g. by a factor of up to 5 by representing the sequence in a partly or non-overlapping set. The advantage of using a completely overlapping set is that it provides a more precise location of any sequence difference, as the mismatch will scan in s consecutive oligonucleotides.

4. ANALYSIS OF AN UNDETERMINED SEQUENCE

The genomes of all free living organisms are larger than a million base pairs and none has yet been sequenced completely. Restriction site mapping reveals only a small part of the sequence, and can detect only a small portion of mutations when used to compare two genomes. More efficient methods for analysing complex sequences are needed to bring the full power of molecular genetics to bear on the many biological problems for which there is no direct access to the gene or genes involved. In many cases, the full sequence of the nucleic acids need not be determined; the important sequences are those which differ between two nucleic acids. To give three examples: the DNA sequences which are different between a wild type organism and one which carries a mutant can lead the way to isolation of the relevant gene; similarly, the sequence differences between a cancer cell and its

normal counterpart can reveal the cause of transformation; and the RNA sequences which differ between two cell types point to the functions which distinguish them. These problems can be opened to molecular analysis by a method which identifies sequence differences. Using the approach outlined here, such differences can be revealed by hybridising the two nucleic acids, for example the genomic DNA of the two genotypes, or the mRNA populations of two cell types to an array of oligonucleotides which represent all possible sequences. Positions in the array which are occupied by one sequence but not by the other show differences in two sequences. This gives the sequence information needed to synthesise probes which can then be used to isolate clones of the sequence involved.

4.1 ASSEMBLING THE SEQUENCE INFORMATION

Sequences can be reconstructed by examining the result of hybridisation to an array. Any oligonucleotide of length s from within a long sequence, overlaps with two others over a length s-1. Starting from each positive oligonucleotide, the array may be examined for the four oligonucleotides to the left and the four to the right that can overlap with a one base displacement. If only one of these four oligonucleotides is found to be positive to the right, then the overlap and the additional base to the right determine s bases in the unknown sequence. The process is repeated in both directions, seeking unique matches with other positive oligonucleotides in the array. Each unique match adds a base to the reconstructed sequence.

4.2 SOME STATISTICS

Any sequence of length N can be broken down to a set of $\sim$ N overlapping sequences s base pairs in length. (For double stranded nucleic acids, the sequence complexity of a sequence of N base pairs is 2N, because the two strands have different sequences,

but for the present purpose, this factor of two is not significant). For oligonucleotides of length s, there are 4^s different sequence combinations. How big should s be to ensure that most oligonucleotides will be represented only once in the sequence to be analysed, of complexity N? For a random sequence the expected number of s-mers which will be present in more than one copy is

$$\mu_{>1} \approx 4^s(1 - e^{-\lambda}(1 + \lambda))$$

where

$$\lambda = (N - s + 1)/4^s$$

For practical reasons it is also useful to know how many sequences are related to any given s-mer by a single base change. Each position can be substituted by one of three bases, there are therefore 3s sequences related to an individual s-mer by a single base change, and the probability that any s-mer in a sequence of N bases is related to any other s-mer in that sequence allowing one substitution is $3s \times N/4^s$. The relative signals of matched and mismatched sequences will then depend on how good the hybridisation conditions are in distinguishing a perfect match from one which differ by a single base. (If 4^s is an order of magnitude greater than N, there should only be a few, 3s/10, related to any oligonucleotide by one base change.) The indications are that the yield of hybrid from the mismatched sequence is a fraction of that formed by the perfect duplex.

For what follows, it is assumed that conditions can be found which allow oligonucleotides which have complements in the probe to be distinguished from those which do not.

4.3 ARRAY FORMAT, CONSTRUCTION AND SIZE

To form an idea of the scale of the arrays needed

to analyse sequences of different complexity it is convenient to think of the array as a square matrix. All sequences of a given length can be represented just once in a matrix constructed by drawing four rows representing the four bases, followed by four similar columns. This produces a 4 x 4 matrix in which each of the 16 squares represents one of the 16 doublets. Four similar matrices, but one quarter the size, are then drawn within each of the original squares. This produces a 16 x 16 matrix containing all 256 tetranucleotide sequences. Repeating this process produces a matrix of any chosen depth, s, with a number of cells equal to 4^s. As discussed above, the choice of s is of great importance, as it determines the complexity of the sequence representation. As discussed below, s also determines the size of the matrix constructed, which must be very big for complex genomes. Finally, the length of the oligonucleotides determines the hybridisation conditions and their discriminating power as hybridisation probes.

s	4^s	Genomes	Side of Matrix (pixel=100 µm)	Number of Sheets of film
8	65536	$4^s \times 10$		
9	262144			
10	1.0×10^6	cosmid	100 mm	1
11	4.2×10^6			
12	1.7×10^7			
13	6.7×10^7	E.coli		
14	2.6×10^8	yeast	1.6 m	9
15	1.1×10^9			
16	4.2×10^9			
17	1.7×10^{10}			
18	6.7×10^{10}	human	25 m	2,500
19	2.7×10^{11}			
20	1.1×10^{12}		100 m	

The table shows the expected scale of the arrays needed to perform the first analysis of a few genomes. The examples were chosen because they are genomes which have either been sequenced by conventional procedures - the cosmid scale -, are in the process of being sequenced - the E. coli scale -, or for which there has been considerable discussion of the magnitude of the problem - the human scale. the table shows that the expected scale of the matrix approach is only a small fraction of the conventional approach. This is readily seen in the area of X-ray film that would be consumed. It is also evident that the time taken for the analysis would be only a small fraction of that needed for gel methods. The "Genomes" column shows the length of random sequence which would fill about 5% of cells in the matrix. This has been determined to be the optimum condition for the first step in the sequencing strategy discussed below. At this size, a high proportion of the positive signals would represent single occurrences of each oligomer, the conditions needed to compare two genomes for sequence differences.

5. REFINEMENT OF AN INCOMPLETE SEQUENCE

Reconstruction of a complex sequence produces a result in which the reconstructed sequence is interrupted at any point where an oligomer that is repeated in the sequence occurs. Some repeats are present as components of long repeating structures which form part of the structural organisation of the DNA, dispersed and tandum repeats in human DNA for example. But when the length of oligonucleotide used in the matrix is smaller than that needed to give totally unique sequence representation, repeats occur by chance. Such repeats are likely to be isolated. That is, the sequences surrounding the repeated oligomers are unrelated to each other. The gaps caused

by these repeats can be removed by extending the sequence to longer oligomers. In principle, those sequences shown to be repeated by the first analysis, using an array representation of all possible oligomers, could be resynthesised with an extension at each end. For each repeated oligomer, there would be 4 x 4 = 16 oligomers in the new matrix. The hybridisation analysis would now be repeated until the sequence was complete. In practice, because the results of a positive signal in the hybridisation may be ambiguous, it may be better to adopt a refinement of the first result by extending all sequences which did not give a clear negative result in the first analysis. An advantage of this approach is that extending the sequence brings mismatches which are close to the ends in the shorter oligomer, closer to the centre in the extended oligomer, increasing the discriminatory power of duplex formation.

5.1 A HYPOTHETICAL ANALYSIS OF THE SEQUENCE OF BACTERIOPHAGE λ DNA

Lambda phage DNA is 48,502 base pairs long. Its sequence has been completely determined, we have treated one strand of this as a test case in a computer simulation of the analysis. The table shows that the appropriate size of oligomer to use for a sequence of this complexity is the 10-mer. With a matrix of 10-mers, the size was 1024 lines square. After "hybridisation" of the lambda 10-mers in the computer, 46,377 cells were positive, 1957 had double occurrences, 75 triple occurrences, and three quadruple occurrences. These 46,377 positive cells represented known sequences, determined from their position in the matrix. Each was extended by four x one base at the 3' end and four x one base at the 5', end to give 16 x 46,377 = 742,032 cells. This extended set reduced the

number of double occurrences to 161, a further 16-fold extension brought the number down to 10, and one more provided a completely overlapped result. Of course, the same end result of a fully overlapped sequence could be achieved starting with a 4^{16} matrix, but the matrix would be 4000 times bigger than the matrix needed to represent all 10-mers, and most of the sequence represented on it would be redundant.

5.2 LAYING DOWN THE MATRIX

The method described here envisages that the matrix will be produced by synthesising oligonucleotides in the cells of an array by laying down the precursors for the four bases in a predetermined pattern, an example of which is described above. Automatic equipment for applying the precursors has yet to be developed, but there are obvious possibilites; it should not be difficult to adapt a pen plotter or other computer-controlled printing device to the purpose. The smaller the pixel size of the array the better, as complex genomes need very large numbers of cells. However, there are limits to how small these can be made. 100 microns would be a fairly comfortable upper limit, but could probably not be achieved on paper for reasons of texture and diffusion. On a smooth impermeable surface, such as glass, it may be possible to achieve a resolution of around 10 microns, for example by using a laser typesetter to preform a solvent repellant grid, and building the oligonucleotides in the exposed regions. One attractive possibility, which allows adaptation of present techniques of oligonucleotide synthesis, is to sinter microporous glass in microscopic patches onto the surface of a glass plate. Laying down very large number of lines or dots could take a long time, if the printing mechanism were slow. However, a low cost ink-

jet printer can print at speeds of about 10,000 spots per second. With this sort of speed, 10^8 spots could be printed in about three hours.

5.3 OLIGONUCLEOTIDE SYNTHESIS

There are several methods of synthesising oligonucleotides. Most methods in current use attach the nucleotides to a solid support of controlled pore size glass (CPG) and are suitable for adaptation to synthesis on a glass surface. Although we know of no description of the direct use of oligonucleotides as hybridisation probes while still attached to the matrix on which they were synthesised, there are reports of the use of oligonucleotides as hybridisation probes on solid supports to which they were attached after synthesis. PCT Application WO 85/01051 describes a method for synthesising oligonucleotides tethered to a CPG column. In an experiment performed by us, CPG was used as the support in an Applied Bio-sytems oligonucleotide synthesiser to synthesise a 13-mer complementary to the left hand cos site of phage lambda. The coupling steps were all close to theoretical yield. The first base was stably attached to the support medium through all the synthesis and deprotection steps by a covalent link.

6. PROBES, HYBRIDISATION AND DETECTION

The yield of oligonucleotides synthesised on microporous glass is about 30 µmol/g. A patch of this material 1 micron thick by 10 microns square would hold $\sim 3 \times 10^{-12}$ umol, equivalent to about 2 g of human

DNA. The hybridisation reaction could therefore be carried out with a very large excess of the bound oligonucleotides over that in the probe. So it should be possible to design a system capable of distinguishing between hybridisation involving single and multiple occurrances of the probe sequence, as yield will be proportional to concentration at all stages in the reaction.

The polynucleotide sequence to be analysed may be of DNA or RNA. To prepare the probe, the polynucleotide may be degraded to form fragments. Preferably it is degraded by a method which is as random as possible, to an average length around the chosen length s of the oligonucleotides on the support, and oligomers of exact length s selected by electrophoresis on a sequencing gel. The probe is then labelled. For example, oligonucleotides of length s may be end labelled. If labelled with ^{32}P, the radioactive yield of any individual s-mer even from total human DNA could be more than 10^4 dpm/mg of total DNA. For detection, only a small fraction of this is needed in a patch 10-100 microns square. This allows hybridisation conditions to be chosen to be close to the Tm of duplexes, which decreases the yield of hybrid and decreases the rate of formation, but increases the discriminating power. Since the bound oligonucleotide is in excess, signal need not be a problem even working close to equilibrium.

Hybridisation conditions can be chosen to be those known to be suitable in standard procedures used to hybridise to filters, but establishing optimum conditions is important. In particular, temperature needs to be controlled closely, preferably to better than $\pm 0.5^{o}C$. Particularly when the chosen length of the oligonucleotide is small, the analysis needs to be

able to distinguish between slight differences of rate and/or extent of hybridisation. The equipment may need to be programmed for differences in base composition between different oligonucleotides. In constructing the array, it may be preferable to partition this into sub-matrices with similar base compositions. This may make it easier to define the Tm which may differ slightly according to the base composition.

Autoradiograpy, especially with ^{32}P causes image degradation which may be a limiting factor determining resolution; the limit for silver halide films is around 25 microns. Obviously some direct detection system would be better. Fluorescent probes are envisaged; given the high concentration of the target oligonucleotides, the low sensitivity of fluorescence may not be a problem.

We have considerable experience of scanning autoradiographic images with a digitising scanner. Our present design is capable of resolution down to 25 microns, which could readily be extended down to less than present application, depending on the quality of the hybridisation reaction, and how good it is at distinguishing absence of a sequence from the presence of one or more. Devices for measuring astronomical plates have an accuracy around 1 μ. Scan speeds are such that a matrix of several million cells can be scanned in a few minutes. Software for the analysis of the data is straight-forward, though the large data sets need a fast computer.

CLAIMS

1. Apparatus for analysing a polynucleotide sequence, comprising a support and attached to a surface therof an array of the whole or a chosen part of a complete set of oligonucleotides of a chosen length, the oligonucleotides being capable of taking part in hybridisation reactions.

2. Apparatus for studying mutations in a known polynucleotide sequence, comprising a support and attached to a surface thereof an array of the whole or a chosen part of a complete set of oligonucleotides of a chosen length comprising the known polynucleotides sequence, the oligonucleotides being capable of taking part in hybridisation reactions.

3. Apparatus as claimed in claim 1 or claim 2, wherein the chosen length is from 8 to 20 nucleotides.

4. Apparatus as claimed in any one of claims 1 to 3, wherein the surface of the support to which the oligonucleotides are attached is of glass.

5. Apparatus as claimed in any one of claims 1 to 4, wherein each oligonucleotide is bound to the support through a covalent link.

6. A method of analysing a polynucleotide sequence, by the use of a support to the surface of which is attached an array of the whole or a chosen part of a complete set of oligonucleotides of a chosen length, which method comprises labelling the polynucleotide sequence or fragment thereof to form labelled material, applying the labelled material under hybridisation conditions to the array, and observing the location on the surface of the label associated with particular members of the set of oligonucleotides.

7. A method according to claim 6, applied to the study of mutations in a known polynucleotide sequence, wherein the array is of the whole or a chosen part of the complete set of oligonucleotides of a chosen length comprising the known polynucleotide sequence.

8. A method according to claim 6, wherein the polynucleotide sequence is randomly degraded to form a mixture of oligomers of the chosen length, the mixture being thereafter labelled to form the labelled material.

9. A method as claimed in claim 8, wherein the oligomers are end labelled with ^{32}P.

10. A method as claimed in any one of claims 6 to 9, wherein the chosen length is from 8 to 20 nucleotides.